



**DEPARTMENT OF EDUCATION**

**[Docket ID ED-2022-IES-0051]**

**Request for Information on the Existence and Use of Large Datasets to Address Education Research Questions**

**AGENCY:** Institute of Education Sciences, Department of Education.

**ACTION:** Request for information.

**SUMMARY:** The National Center for Education Research (NCER), a center within the U.S. Department of Education's Institute of Education Sciences, funds and coordinates high-quality, innovative research that addresses the biggest challenges facing education in the 21st century. Through this request for information (RFI), NCER seeks public input to help us identify existing large datasets that may be useful for research and to understand the challenges and limitations that may affect access and their value for research.

**DATES:** We must receive your comments by [INSERT DATE 30 DAYS AFTER DATE OF PUBLICATION IN THE FEDERAL REGISTER].

**ADDRESSES:** Comments must be submitted via the Federal eRulemaking Portal at [regulations.gov](https://www.regulations.gov). However, if you require an accommodation or cannot otherwise submit your comments via [regulations.gov](https://www.regulations.gov), please contact the program contact person listed under FOR FURTHER INFORMATION CONTACT. The Department will not accept comments by fax or

by email. To ensure that the Department does not receive duplicate copies, please submit your comments only once. Additionally, please include the Docket ID at the top of your comments.

The Department strongly encourages you to submit any comments or attachments in Microsoft Word format. If you must submit a comment in Adobe Portable Document Format (PDF), the Department strongly encourages you to convert the PDF to "print-to-PDF" format, or to use some other commonly used searchable text format. Please do not submit the PDF in a scanned format. Using a print-to-PDF format allows the Department to electronically search and copy certain portions of your submissions to assist in the process.

Federal eRulemaking Portal: Go to [www.regulations.gov](http://www.regulations.gov) to submit your comments electronically. Information on using Regulations.gov, including instructions for accessing agency documents, submitting comments, and viewing the docket, is available on the site under the "FAQ" tab.

Privacy Note: The Department's policy for comments received from members of the public is to make these submissions available for public viewing in their entirety on the Federal eRulemaking Portal at [www.regulations.gov](http://www.regulations.gov). Therefore, commenters should be careful to include in their comments only information that they wish to make publicly available. We encourage, but do not require, that each

respondent include their name, title, institution or affiliation, and the name, title, mailing and email addresses, and telephone number of a contact person for the institution or affiliation, if any.

FOR FURTHER INFORMATION CONTACT: Erin Higgins, Program Officer, National Center for Education Research, Institute of Education Sciences, U.S Department of Education, 400 Maryland Avenue, SW, Washington, DC 20202-7240. Telephone: (202) 706-8509. You may also email your questions to Erin.Higgins@ed.gov, but as described above, comments must be submitted via the Federal eRulemaking Portal at regulations.gov.

If you are deaf, hard of hearing, or have a speech disability and wish to access telecommunications relay services, please dial 7-1-1.

**SUPPLEMENTARY INFORMATION:**

Background:

The number of large education-related datasets is growing, and we have new opportunities to leverage these data to address critical questions of policy and practice. For example, State longitudinal data systems (SLDS) can support research on the questions that State agencies have about a specific education issue, program, or policy. SLDSs have the potential to support lower-cost, faster research by avoiding the need for costly primary data collection. Similarly, education technologies generate

large amounts of data that--after ensuring students' privacy is protected--can potentially provide valuable insights about learning. Despite the large amount of raw data collected by these technologies, there are legal, practical, and methodological barriers to conducting research that leverages these types of datasets to understand and improve students' education outcomes. Education researchers seeking to conduct studies using these datasets confront challenges related to the validity of data elements and the logistics of data access in ways that protect students' privacy, consistent with local, State, and Federal law. Researchers face significant barriers and costs to access these datasets, which leads to only a small number of education studies with large sample sizes, despite the known advantages of these types of studies.

There are examples of the potential insights to be gained from these data, and the fields of educational data mining and learning analytics have developed methods and insights for working with large datasets. For example, researchers have analyzed data collected in the digital administration of NAEP, which has led to insights into multiple aspects of student test-taking strategies.<sup>12</sup>

---

<sup>1</sup> Arslan, B., Gong, T., Feng, G., Agard, C., & Keehner, M. (2021, June 8). *Going beyond scores: Understanding fourth-graders' scientific inquiry practices with process data*. [Paper presentation]. The 2021 Virtual Annual Meeting of the National Council on Measurement in Education.

<sup>2</sup> Wang, N. & Circi, R. (2020, August). *Revisiting Omit and Not-Reached Scoring Rule using NAEP Process Data*. In J. Weeks (Chair). *Diving into*

Data privacy is central to the ethical conduct of research. Any plans to leverage the large amounts of data that are being collected through education technology, State longitudinal data systems, and other sources must be designed to minimize the risk of disclosure in order to protect the privacy of students.

Through this RFI, we seek public comment to help us identify existing large datasets, especially those that are generated using education technology, that may be useful for research; identify best practices for creating new, large datasets that are valuable for research; understand the challenges and limitations that may impact data access; and develop and implement plans to protect students' privacy.

This is a request for information only. This RFI is not a request for proposals (RFP) or a promise to issue an RFP or a notice inviting applications. This RFI does not commit the Department to contract for any supply or service whatsoever. Further, we are not seeking proposals and will not accept unsolicited proposals. The Department will not pay for any information or administrative costs that you may incur in responding to this RFI. The documents and information submitted in response to this RFI will not be returned.

We will review every comment, and the comments in response to this RFI will be publicly available on the Federal eRulemaking Portal at [www.regulations.gov](http://www.regulations.gov). Please note that IES will not directly respond to comments.

Solicitation of Comments:

We invite stakeholders who are aware of large datasets relevant to education and learning, especially those generated through education technology; stakeholders who have perspectives on the value of these datasets for education research; and stakeholders who are aware of challenges and limitations to both access and use of large datasets to share responses to the following questions in their comments:

(1) What public or restricted use education-related datasets are available for training students in data mining/machine learning methods? What training needs are not being met by the datasets that are currently available?

(2) What open or restricted use education-related datasets are available to train new artificial intelligence models or to test hypotheses using data mining/machine learning methods? What research needs are not being met by the datasets that are currently available?

(3) What work do researchers need to do to access, and then explore the quality of, an existing dataset before conducting research with it? What aspects of this work could be reduced or conducted just once so that future

researchers can reduce the time needed to complete a research project?

(4) How do researchers determine the validity of data elements within previously collected datasets? What challenges are frequently encountered related to how those data align to constructs of interest?

(5) What are promising approaches to testing and improving the validity of metrics within large datasets, especially those datasets that are developed through interactions with education technology?

(6) How likely is it that existing datasets, especially those that come out of education technology, contain data that are valuable for researchers and of sufficient quality that research could be conducted with a high amount of rigor?

(7) To what extent do existing datasets capture enough information to address research questions related to diversity, equity, inclusion, and accessibility? What additional data should be collected to address these questions?

(8) What are the best practices for creating new datasets or linking existing datasets and sharing them with researchers (open or restricted use) while prioritizing the privacy of individuals and adhering to local, State, and Federal laws? What barriers and limitations exist?

(9) What role can IES play in developing

infrastructure that supports the use of large-scale datasets for education research?

Accessible Format: By request to the program contact person listed under FOR FURTHER INFORMATION CONTACT, individuals with disabilities can obtain this document in an accessible format. The Department will provide the requestor with an accessible format that may include Rich Text Format (RTF) or text format (txt), a thumb drive, an MP3 file, braille, large print, audiotape, or compact disc, or other accessible format.

Electronic Access to This Document: The official version of this document is the document published in the *Federal Register*. You may access the official edition of the *Federal Register* and the Code of Federal Regulations at [www.govinfo.gov](http://www.govinfo.gov). At this site you can view this document, as well as all other documents of this Department published in the *Federal Register*, in text or Portable Document Format (PDF). To use PDF you must have Adobe Acrobat Reader, which is available free at the site.

You may also access documents of the Department published in the *Federal Register* by using the article search feature at [www.federalregister.gov](http://www.federalregister.gov). Specifically, through the advanced search feature at this site, you can



limit your search to documents published by the Department.

**Mark Schneider,**  
*Director,*  
*Institute of Education Sciences.*

[FR Doc. 2022-09239 Filed: 4/28/2022 8:45 am; Publication Date: 4/29/2022]